

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)

24-06-1999

2. REPORT TYPE

Final

3. DATES COVERED (From - To)

4. TITLE AND SUBTITLE

Applying Non-Parametric Methods for Calculating and Analyzing Inventory Reorder Points to Military Supply Systems with Stochastic Demands and Fixed or Variable Lead Times

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Ronald D. Fricker, Jr. and Capt. Christopher A. Goodhart, USMC

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The RAND Corporation  
1700 Main Street  
Santa Monica, CA 90407

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

The RAND Corporation  
1700 Main Street  
Santa Monica, CA 90407

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

20000804 188

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This paper develops and applies a nonparametric bootstrap methodology for setting inventory reorder points and a simple inequality for identifying existing reorder points that are unreasonably high. We demonstrate that an empirically-based bootstrap method is both feasible and calculable for large inventories by applying it to the 1st Marine Expeditionary Force General Account, an inventory consisting of \$20-30 million of stock for 10-20,000 different types of items. Further, we show that the bootstrap methodology works significantly better than the existing methodology based on mean days of supply. In fact, we demonstrate performance equivalent to the existing system with a reduced inventory at one-half to one-third the cost; conversely, we demonstrate significant improvement in fill rates and other inventory performance measures for an inventory of the same cost.

15. SUBJECT TERMS

Logistics, Nonparametric, Inventory, Lead Time Demand, Supply Systems, Markov's Inequality, Bootstrap, Military, United States Marine Corps

16. SECURITY CLASSIFICATION OF:

a. REPORT  
UNCLASSIFIED

b. ABSTRACT  
UNCLASSIFIED

c. THIS PAGE  
UNCLASSIFIED

17. LIMITATION  
OF ABSTRACT

UNCLASSIFIED  
UNLIMITED

18. NUMBER  
OF PAGES

25

19a. NAME OF RESPONSIBLE PERSON  
Dr. Ronald D. Fricker, Jr.

19b. TELEPHONE NUMBER (include area code)  
310-393-0411 x6102

**Applying Non-Parametric Methods for Calculating and Analyzing  
Inventory Reorder Points to Military Supply Systems  
with Stochastic Demands and Fixed or Variable Lead Times**

Dr. Ronald D. Fricker, Jr.,<sup>†</sup>  
and  
Capt. Christopher A. Goodhart, USMC\*

<sup>†</sup>The RAND Corporation, 1700 Main Street, Santa Monica, CA 90407  
\*DC/S Installations and Logistics (LX), 2 Navy Annex, Washington, DC 20380-1775

**67th MORS Symposium**  
Working Group Number 19  
24 June 1999

**ABSTRACT**

This paper develops and applies a nonparametric bootstrap methodology for setting inventory reorder points and a simple inequality for identifying existing reorder points that are unreasonably high. We demonstrate that an empirically-based bootstrap method is both feasible and calculable for large inventories by applying it to the 1st Marine Expeditionary Force General Account, an inventory consisting of \$20-30 million of stock for 10-20,000 different types of items. Further, we show that the bootstrap methodology works significantly better than the existing methodology based on mean days of supply. In fact, we demonstrate performance equivalent to the existing system with a reduced inventory at one-half to one-third the cost; conversely, we demonstrate significant improvement in fill rates and other inventory performance measures for an inventory of the same cost.

**INTRODUCTION**

One almost universal function of any military supply system is to warehouse goods in anticipation of customer demands. Similar requirements exist in portions of the commercial sector and other operations where the advantages of holding stock locally offset the cost of maintaining and managing the stock. A particular advantage to the military is the ability to satisfy a customer demand nearly immediately with on-hand stock, using the local system as a buffer between the customer and the vagaries of the rest of the supply system. Maintaining local stocks requires determining which items to stock and in what quantities. In peacetime, overstocking ties up funds in assets that turn over very slowly and in wartime results in masses of unused material that must be transported or left behind. Understocking defeats the purpose of maintaining local inventory when items are not available and can have a large impact on wartime sustainability.

For a particular stocked item, the *reorder point* (ROP) specifies the stock level at which a replenishment order is placed. The ROP is usually set to provide an acceptably low risk of stock-out between the time when a replenishment order is placed and subsequently received (which, in a system with automatic reordering, is the *lead time*). The quantity of demands occurring in a lead time is referred to as the *lead time demand* (LTD). Setting the ROP is a trade-off between the cost of incurring one or more stock-outs and the cost of holding additional stock (commonly called *safety stock*) to cover the possibility of unanticipated lead time demand. In the military, the penalty for such a stock-out is delay time during which the customer must wait while the part is ordered and delivered from the next higher echelon of supply. In the commercial world, stock-outs may represent lost sales as customers go elsewhere to satisfy their demand.

Classic inventory theory models the distribution of an item's demands parametrically, making particular assumptions about the variability of demands and lead times when setting the reorder point. Poisson, normal, and negative binomial distributions are routinely used. Such distributions allow simple calculation of the demand's variance so that the ROP can be specified in terms of a mean demand plus some multiple of the standard deviation. If the distributional assumptions are appropriate then this approach works well. However, such distributional assumptions rarely apply to United States Marine Corps (USMC) data.

We address this problem with an empirically-based resampling strategy, which uses bootstrapping techniques on historical demand data, to estimate the lead time demand distribution. From this, quantiles for a given probability of stock-out

can then be used to set ROPs. We also derive a simple inequality from Markov's Inequality, which uses only the mean demand rate and mean lead time, to identify whether existing ROPs are set too high. We demonstrate the utility of these tools on actual USMC supply data.

### Related Research

We focus on reducing quantities of stock (depth) by establishing reasonable reorder points while maintaining control of the service level provided to the customer, instead of attempting to minimize a cost function. We seek to do this from an analytically sound basis while relying on an extremely selective set of assumptions. This is because we do not have reliable estimates of standard cost parameters, and can assemble only limited information relating to the distribution of lead time demand (LTD). Fortunately, earlier work provides suitable precedent for approximate, nonparametric or "distribution-free" approaches to establishing inventory reorder points without reliance on restrictive assumptions about demand or lead time, or on cost parameters that can be misleading or hard to obtain in practice (see, e.g., Selen and Wood, 1987).

Many of these approaches have been prompted by concern with appropriateness of ascribing the normal distribution to LTD. Tyworth and O'Neill (1997) review the salient arguments, and examine the errors associated with use of the normal distribution to model LTD for the case of fast-moving finished goods inventories when the coefficient of variation is at or below 0.4. They conclude that normal theory adequately models cost and service characteristics in that setting. Eppen and Martin (1988) demonstrate instances in which traditional normal-theory techniques can produce unacceptable reorder points.

They illustrate that utilizing quantiles of the LTD distribution, constructed from separate demand and lead time distributions, can produce more accurate reorder points, and provide a technique for establishing reorder points when no distributional information is available but a demand forecasting system is in place. Tijms and Groenevelt (1984) develop tractable two-moment approximations for reorder points in  $(s, S)$  inventory systems in the presence of service level constraints. Approximation accuracy decreases with increases in the LTD coefficient of variation and for high desired service levels; the approximations also rely on relatively accurate calculation of the first two moments of the distribution of LTD.

Azoury and Brill (1992) determine the stationary distribution of the net inventory process in an  $(s, S)$  inventory system as a function of  $s$  and  $S$ . They assume demands arrive according to a Poisson process and, in the most general cases, that demand quantities and lead times are i.i.d. following general distributions. The authors suggest that if these distribution functions are not known, the family of generalized hyperexponential distributions can represent LTD suitably. Further, under the assumption that LTD will exceed  $S - s$  with negligible probability, their approximations can be simplified substantially. The results of Azoury and Brill are appealing, because possession of the stationary distribution of net inventory provides the capability of relating policy parameters – reorder point and requisitioning objective – to a number of measures of effectiveness, among them a number of service measures.

The “min-max” approach is a conservative technique potentially appropriate when faced with high stock out or disposal costs, or with the requirement to provide a consistently high service level. Gallego (1992), Gallego and Moon (1993),

Moon and Gallego (1994), and Ouyang and Wu (1997) apply a min-max distribution free technique to a variety of stochastic inventory problems. Gallego (1992) studies the continuous review  $(r, Q)$  inventory problem seeking values of  $r$  and  $Q$  that minimize long run average total cost when only the mean and variance of LTD are known. The two-phased approach first determines the expected backorders incurred corresponding to the worst possible distribution of lead time demand possessing the given mean and variance, and then minimizes the corresponding long run average total cost with respect to  $r$  and  $Q$ . Moon and Gallego (1994) extend Gallego’s analysis to a periodic review model with a mixture of backorders and lost sales; Ouyang and Wu (1997) allow mixed backorders and lost sales while adding a cost for lead time reduction and implementing a service level constraint. They also provide a tighter bound on stock out probability than we do here, though they require the first two moments of LTD to calculate it.

Several distribution-free techniques for quantile estimation based on order statistics appear in the literature. In an inventory setting, Lordahl and Bookbinder (1994) interpolate between closest order statistics of observed lead time demands to estimate quantiles of the LTD distribution. They subsequently compare cost and service performance of the resulting reorder point estimates against corresponding performance of reorder points developed by assuming LTD is normally distributed. Harrell and Davis (1982) develop a quantile estimator based on a linear combination of all order statistics and provide examples to show that it is generally more efficient than traditional simple interpolation. Both papers indicate the conditions under which the selected alternate methodology outperforms their technique, Lordahl and Bookbinder’s in

terms of service and cost, and Harrell and Davis' in terms of estimator efficiency.

Bookbinder and Lordahl (1989) and Wang and Rao (1992) address reorder point determination via a bootstrap procedure. In Bookbinder and Lordahl (1989), the authors compare reorder points determined by bootstrapping from observed lead time demands with traditional normal theory across many different distributional forms of LTD, incorporating cost and service level measures in the analysis. They conclude in the majority of cases that the bootstrap performs as well or better than the normal approach, and find it to be preferable when LTD is bimodal, highly skewed, or otherwise not well described by standard distributions. Wang and Rao (1992) extend the comparison and further develop bootstrap reorder point estimates in the specific case of correlated demand. Wang and Rao (1992) bootstrap from observed LTDs and discuss the application to LTDs generated via Monte Carlo simulation.

Marine Corps consumable item retail inventory systems possess characteristics that frustrate attempts to implement many of these techniques. The principal cause of difficulty is that these inventories contain large numbers of very different types of items (usually greater than 10,000) with widely varying demand patterns. Available data represent a continuum between fast-moving, commercially available products and specialized military equipment with much more sporadic demand and longer lead times. It is not feasible to manage each item individually, and specific assumptions required for implementation of many existing techniques do not apply to all, or even most, items. For example, median LTD coefficient of variation is 4.0; only 88 LTD distributions out of 32,650 examined pass the Shapiro-Wilk test for normality; and 90% of the items' LTD distributions

have skewness greater than or equal to 2 (a normal distribution has skewness 0).

Even under conditions that depart from the recommended criteria for coefficient of variation and normality, the bootstrap performed well in Bookbinder and Lordahl's evaluations. We employed a similar bootstrap technique because it accommodates virtually any distributional form, because it provided sufficiently accurate estimates of the ROPs for our requirements, and because it could ultimately be implemented in the field within an algorithm that did not require sophisticated user input. We show that our application of the methodology to actual Marine Corps data results in an inventory that costs one-half to one-third of the current inventory yet performs just as well. We also show that using our bootstrap methodology to define an inventory with a value of the existing inventory results in significantly improved system performance as measured by multiple inventory performance metrics.

The paper is organized as follows. In the "Background" section we provide information about the Marine Corps' supply system and some basic inventory theory, including how the USMC currently determines inventory ROPs and the metrics it uses to measure supply performance. In the "Using the Bootstrap to Calculate Reorder Points" section we describe how to apply our bootstrap methodology to calculate ROPs. We follow this with a section on the "Evaluation of Reorder Point Performance with Actual Marine Corps Data" which: (1) evaluates our bootstrap methodology versus the existing USMC method and other techniques recommended in the literature for ROP calculation; (2) presents a service-dependent upper bound on the reorder point based on Markov's Inequality; and, (3) illustrates bootstrap ROP performance using real demand data.



"Evaluating the Overall Effect of the Bootstrap ROPs" describes the results of a more detailed experiment, incorporating bootstrap reorder points into a realistic evaluation of overall inventory performance. We conclude in the "Summary" section with a discussion and recommendations. A glossary of acronyms is provided at the end of the paper.

## BACKGROUND

Methods of managing local stock vary by military service (and commercial organization) and may be comprised of one or more echelons of supply. Here we only consider the problem from the perspective of local ("retail") inventory established to serve a set of local customers; it consists of a single set of stocks which are periodically replenished as necessary from an unspecified higher echelon or echelons of supply. This is essentially the arrangement at each Marine Expeditionary Force's (MEF) *General Account*. The General Account is a local inventory, physically located with the MEF, consisting primarily of repair parts but also including many other types of consumable material. The dollar value of the on-hand stock in a typical MEF General Account ranges between 20 to 30 million dollars, for which the managing organization has an annual budget of the same order of magnitude that is allocated incrementally on a quarterly basis. The General Account may stock between ten and twenty thousand individual types of items ("lines") identified by *national stock number* (NSN), the military equivalent of the commercial stock keeping unit (s.k.u.), and the inventory may experience anywhere from two to three hundred thousand requisitions a year. A *requisition* is a request by a specific customer for one or more units of a particular item; thus the inventory

experiences a total number of demands far in excess of the number of requisitions.

Establishing a local inventory requires determining both which items to stock and the amount of each item to stock. The Marine Corps' decision to stock an item at any particular General Account depends on several factors. If there is no advance knowledge of special need for an item, the decision to stock it is based on local usage and the item's *Combat Essentiality Code*. Combat essential items not previously stocked cannot normally be stocked unless the General Account has issued three or more of the item in the previous 12 months; non-combat essential items require a minimum of six demands in 12 months (United States Marine Corps *Intermediate-Level Supply Management Policy Manual*, 1992). Decisions to stock particular items on criteria other than minimum usage—decisions the Marine Corps calls "nondemand-supported"—are made on an exception basis and reviewed, at least annually, apart from standard retention criteria.

When an inventory manager at a General Account decides, based on the preceding criteria, to stock an item, its subsequent management follows a standard ( $s, S$ ) inventory policy. The reorder point  $s$  is the ROP; the order-up-to level  $S$  is called the *requisitioning objective* (RO). The Marine Corps refers to the difference  $S - s$  as the *operating level*, a quantity it establishes separately from the ROP. Inventory position—on-hand plus due-in minus backorders—triggers a replenishment order when it reaches the reorder point (a computer provides a recommendation to make the buy, but personnel at the General Account must submit the replenishment order manually).

The remainder of our discussion centers principally on determination of the

reorder point. This determination requires measurement of demand and lead time as in many inventory systems. We refer to *lead time* as the elapsed time between when the inventory position reaches the ROP and when the subsequent replenishment order is available to satisfy future demands. The *order and ship time* (OST) is a component of the lead time and is usually considered to be the time from when the order is placed with the next higher echelon to when it is received locally. If the replenishment orders are placed automatically when the ROP is reached, and if the time to re-stock the local inventory is negligible, then the lead time is equal to the OST.

#### Current Methodology

The Marine Corps currently utilizes *days of supply*, a demand-dependent unit of measure, as the primary way to set the requisitioning objective and ROP. One day of supply is the mean daily usage of an item over some past time period, such as the last year. At the 1<sup>st</sup> Marine Expeditionary Force (I MEF) General Account, the ROP is defined as the sum of

- safety stock of either 15 or 30 days of supply depending on the item's Combat Essentiality Code: 30 days of supply for items coded as combat essential, 15 days of supply otherwise, and
- lead time stock based on the estimated OST (converted to days of supply) for each item.

Thus, for a hypothetical combat essential item with a historical mean demand of 3 per day, and an OST of 11 days, the ROP would be 123 (30 days of supply x 3 units/day + 11 days of supply x 3 units/day). The requisitioning objective for an item is defined as the ROP plus an operating level of 60 days of supply. For the hypothetical item, the requisitioning objective would be set at 303 units. The Army uses essentially

the same methodology, differing in the specified days of supply for the operating level and safety stock and in the methodology for estimating item lead times.

There are a number of weaknesses with this methodology. First, and most importantly, it eliminates information about demand variability contained in the data by only using means. For example, an item that deterministically has one demand for one unit a day for 365 days has exactly the same days of supply, and thus the same requisitioning objective and ROP, as another item that only has a single demand on one day for 365 units and no demands in the other 364 days. Second, calculating lead times for individual items that have infrequent replenishment orders using the observed OSTs may result in poor lead time estimation because the sample sizes are small. Finally, for organizations that do not use automatic replenishment ordering, the OST does not represent the complete lead time. This is because the time from when an item's inventory position exceeds the ROP until the time the replenishment order is placed is not included in the OST. If this time period is significant then the safety level does not reflect the entire time the item was at or below the ROP and at risk of stock-out.

#### Performance Metrics

The usual metric for measuring inventory performance is *fill rate*, which is the ratio of demands filled out of local inventory to the total number of demands presented over some time period. This measure of performance can be subdivided into two others: The *accommodation rate* is the ratio of the number of demands for items locally stocked to the total number of demands, and the *satisfaction rate* is the ratio of the number of demanded items filled out of inventory among the set of demands

for items which are locally stocked to the total number of demands for items which are locally stocked. That is, the satisfaction rate is a conditional fill rate for only those demands which are stocked locally, and the overall fill rate is the product of the satisfaction and accommodation rates.

Fill rate alone is not a sufficient measure to characterize how well an inventory supports its customers because (a) it treats all items equally and (b) it does not capture the effects on the customer who needs a set of items. For example, the Marine Corps' General Accounts provide repair parts for broken equipment. It is common for a mechanic to need a complete set of parts to finish (and sometimes even to start) a repair. Thus even with an 80 or 90 percent fill rate it may often be true that repairs are held up because one or more parts must be ordered from the next higher echelon of supply. The result is that if one part cannot be filled out of the local inventory then it might just as well have been that none of the parts were filled since the entire repair is likely to have to wait for the final part to arrive anyway. The Marine Corps measures this type of inventory performance via the *Equipment Repair Order (ERO) fill rate* metric. An ERO is essentially a work order for the repair of a piece of equipment, a portion of which lists the required repair parts. The ERO fill rate is the fraction of important EROs for which all of the high priority parts were immediately available from the local General Account inventory. (Specific details regarding what constitutes an "important" ERO, a "high priority" part, and how "immediately available" is defined can be found in Fricker and Robbins, 1999.) Typical General Account fill rates are between 50 and 70 percent; typical ERO fill rates are between 40 and 60 percent.

## USING THE BOOTSTRAP TO CALCULATE REORDER POINTS

As discussed in any standard inventory theory text (see, for example, Tersine, 1994, or Arrow, Karlin, and Scarf, 1958), there are three basic inventory problems, listed below in order of increasing complexity:

1. Fixed demand and lead time,
2. Stochastic demand and fixed lead time, and
3. Stochastic lead time and demand.

Also, some production planning models may assume deterministic demand and random lead time, which results (like cases 2 and 3) in the definition of a lead time demand distribution. The first case will not be considered further as it is trivial to solve arithmetically (though we note that the bootstrap methodology to be described applies equally well to this case as to the more complex, stochastic cases). For the second and third cases, let  $t$  denote time for some basic unit (days, weeks, etc.) and let  $D_t$  denote the number of items requested (the demand) during period  $t$ . Let  $L$  denote the lead time (in the same units as  $t$ ) for an item's replenishment order and let  $\alpha$  denote the probability that a stock-out occurs during the lead time. Assume that the demands  $D_t$  and the lead times  $L$  are both i.i.d. from unspecified distributions, and  $D_t$  and  $L$  are independent. Let  $X(J, L) = \sum_{t=J+1}^{J+L} D_t$  represent

the cumulative demands occurring in  $L$  successive time periods after some time  $J$  when the inventory position is first at or below the ROP. Denote  $P\{X(J, L) \leq x\}$  as  $F(x)$ , the distribution function of LTD. To achieve a particular probability of stock out when the inventory is at the ROP, set  $x = F^{-1}(1 - \alpha)$  where  $x$  is the chosen ROP and then  $P\{X(J, L) > x\} = \alpha$ . We refer to  $\alpha$  as the *risk* of stock-out; the quantity  $100(1 - \alpha)$  is often referred to as the *service level*.



Both Bookbinder and Lordahl (1989) and Wang and Rao (1992) bootstrap observations from the LTD, assuming that actual observations for  $m$  lead times are available,  $\{X_1, \dots, X_m\}$ . We do not use this approach for two reasons: (1) Our data do not contain information about the inventory position, so we cannot identify the actual lead time periods, and (2) under the assumption that demands are independent of inventory position, only using data from the lead times ignores the majority of demand data. That is, if demands are independent of inventory position, then for any particular  $L$  it follows that  $X(J, L) \stackrel{d}{=} X(j, L)$  for  $j=1, 2, \dots$ , and so we need not restrict the data to only the lead time periods.

Under these assumptions, for a fixed lead time  $L$ , the bootstrap (Efron and Tibshirani, 1993) can be used to create an empirical distribution  $\hat{F}$  as follows. Suppose that for each item  $i$  a set of historical demands exists for  $N$  periods:  $D_1^i, \dots, D_N^i$ ,  $N\tau L$ . For a fixed lead time of  $L$  periods, the bootstrap methodology is applied by randomly drawing with replacement  $L$  observations from  $\{D_1^i, \dots, D_N^i\}$  many, say  $M$ , times. The bootstrap observations are denoted  $D_{(1)}^i, \dots, D_{(L)}^i$ , with the iteration subscript suppressed for clarity. For each iteration, then, the bootstrap demand quantity for a lead time period is calculated as  $\hat{X}^i = \sum_{j=1}^L D_{(j)}^i$ . The empirical distribution  $\hat{F}_i$  is calculated from the  $M$  bootstrap statistics,  $\mathbf{X}^i = \{\hat{X}_1^i, \dots, \hat{X}_M^i\}$ , and the ROP is then set choosing the quantile corresponding to the desired risk  $\alpha$ ,  $ROP_i = \hat{F}_i^{-1}(1 - \alpha)$ .

For the variable lead time problem, the bootstrap methodology is further adapted by

first randomly sampling from an empirical distribution of lead times and then resampling the number of demand periods corresponding to the chosen lead time. For example, given an empirical lead time distribution for item  $i$ , a lead time is randomly selected; say  $l$  was chosen, then  $l$  periods are randomly drawn from the historical data, from which  $\hat{X}^i(l)$  is calculated. As before this process is repeated  $M$  times, where at each iteration a different lead time is randomly drawn from the historical lead time distribution.

Since the empirical distribution  $\hat{F}$  is discrete, some method will generally have to be employed to select the quantile which gives the risk closest (in some sense) to the specified risk. Efron and Tibshirani (1993) describe various methods for selecting quantiles which involve ranking the bootstrap observations and then choosing the  $\alpha \cdot M$  observation (with rules for handling  $\alpha \cdot M$  not integer). In our work here we set the ROP to the value in the empirical distribution with the risk closest to the desired risk in terms of Euclidean distance.

That is,  $ROP_i = \{x : \min_{\forall x \in \mathbf{X}^i} |\hat{F}_i(x) - \alpha|\}$ , where  $\hat{F}$  is the cumulative tail function of  $\hat{F}$  (i.e.,  $\hat{F}(z) = 1 - \hat{F}(z)$ ). This is a more conservative approach than ranking, in the sense that ranking will always choose an ROP greater than or equal to our ROP, but our method generally requires less computer disk storage space because fewer than  $M$  observations can be stored when there are tied bootstrap observations. Other rules can also be used to choose the ROP; for example, we will describe an interpolation scheme by Lordahl and Bookbinder (1994) in "Comparison of Bootstrap, Order-Statistics, and Normal Approximation Techniques" subsection.

### Intuition Behind the Methodology

Some intuition is in order for why our approach works and is desirable. First, consider a simple, ideal case in which the lead time for an item is deterministically one day and a large amount of historical demand data is available. Then in order to evaluate the risk of stock-out for setting  $ROP=x$ , one would simply extract from the data all the demands for each day after the inventory position reached or exceeded the ROP and count the fraction of times out of the total number of cases that the next day's demands exceeded  $x$ . This fraction would be a good estimate of the risk under the assumption that the distribution of future demands is similar to the distribution of past demands; the larger the amount of past data and the more similar the past and future demand distributions the better the risk estimate. This is the motivation for constructing an empirical distribution of demands that occurred in the lead time after a replenishment buy order was placed, and it is the motivation for Bookbinder and Lordahl (1989) and Wang and Rao (1992) to bootstrap directly from the LTD. However, under the assumption that inventory position and demands are independent we need not restrict the data only to specific lead time periods, so we can generalize in this simple example by constructing the empirical distribution using all of the daily demands. Such an independence assumption is generally very reasonable to make for military supply systems, since demands are made without any knowledge of inventory position.

For lead times in excess of one day, we can further generalize by starting at the earliest day for which data is available and taking blocks of days sequentially that correspond to the lead time. This approach would allow us to use all of the demand

data, rather than only the lead time data which is (typically) a small subset of the demand data. The difficulty with historical data, even when using all of the demand data, is that it is generally insufficient to construct a useful empirical distribution. For example, for the Marine Corps we only have two years of historical data; average lead times of 20 to 30 days makes for too few LTD observations from which to construct a useful empirical distribution. Furthermore, even if a significant amount of historical demand data were available, the older data could be suspect in terms of distributional stationarity, so that one would probably want to limit the amount of older data utilized. Using the assumption of independent daily demands and independence between lead times and demands, we overcome this obstacle by resampling from the restricted historical data, bootstrapping "alternate" LTD observations which we use to capture the uncertainty related to observing limited demand data. From these alternate observations, we then simply create an empirical distribution just like in the previous simple example.

Under the assumptions of independent daily demands and independence between lead times and demands, the methodology we have described is essentially a simple Monte Carlo LTD distribution estimation scheme. That is, in this sense, one can think about our approach as a convenient computational method to estimate the LTD distribution from the empirical distributions of daily demands and lead times. The connection of this scheme to the usual bootstrap is not obvious, but it is an important connection for generalizing to more complicated data structures and inventory situations, and for calculating a measure of variability of the ROP point estimate.

## Generalizing Under the Bootstrap

### Framework

Our application of the bootstrap might seem to differ from its more common use in estimating the standard error of a mean. To illustrate the more usual application, consider a set of observations  $Y_1, Y_2, \dots, Y_n$  and their mean,  $\bar{Y} = \frac{1}{n} \sum_{k=1}^n Y_k$ . In order to

estimate the standard error of  $\bar{Y}$ , one applies the bootstrap by resampling with replacement  $M$  sets of  $n$  samples from  $Y_1, Y_2, \dots, Y_n$ :  $Y_{(1,k)}, Y_{(2,k)}, \dots, Y_{(n,k)}$ ,  $k=1, \dots, M$ . From these resamples,  $\bar{Y}_{(k)} = \frac{1}{n} \sum_{j=1}^n Y_{(j,k)}$ , is calculated for each  $k$ , and then  $s.e.(\bar{Y})$  is estimated,

$$s.e.(\bar{Y}) = \left[ \frac{1}{M-1} \sum_{k=1}^M (\bar{Y}_{(k)} - \bar{Y}_{(.)})^2 \right]^{1/2},$$

where  $\bar{Y}_{(.)} = \frac{1}{M} \sum_{k=1}^M \bar{Y}_{(k)}$ . Depending on the statistic being estimated,  $\bar{Y}_{(.)}$  may be used as the point estimate, or some other more standard statistic. In this example,  $\bar{Y}$  would probably be used as the point estimate of the mean, and the bootstrap standard error,  $s.e.(\bar{Y})$ , as the measure of the point estimate's variability.

The key ideas in the bootstrap are (1) resampling  $n$  observations with replacement from the original  $n$ , and (2) thinking of these sets of bootstrap observations as "new" data that are representative of the underlying probability distribution that generated the original data. In our application, a computational methodology equivalent to the one we described in the first part of this section is as follows:

1. Resample with replacement  $N$  observations from  $\{D_1^i, \dots, D_N^i\}$  to

generate the bootstrap observations  $D_{(1)}^i, \dots, D_{(N)}^i$ .

2. Then generate a lead time observation  $l$  (stochastically or deterministically) and choose  $l$  days randomly, with replacement, from  $D_{(1)}^i, \dots, D_{(N)}^i$ .
3. Repeat step two  $M$  times, calculate  $\hat{F}_i$  from  $\{\hat{X}_1^i, \dots, \hat{X}_M^i\}$ , and then choose the quantile corresponding to the desired risk  $\alpha$ ,  $ROP_i = \hat{F}_i^{-1}(1 - \alpha)$ .

This represents one bootstrap ROP for item  $i$ , which we could use as the point estimate of the actual distribution quantile. To calculate the standard error for this point estimate, we would repeat steps 1-3 many times and use those replications to compute the standard error in the usual way described above.

In order to see the equivalence between steps 1-3 and the previous Monte Carlo-like description, note that drawing a random sample (with replacement) of lead time demands in step 2 from the bootstrap distribution, which itself was drawn randomly with replacement in step 1, is exactly the same as drawing a random lead time number of demands from the original set of demands. That is, the first two steps simply collapse into one, in which  $l$  demands are randomly drawn with replacement from the original set of  $N$  daily demands. However, under autocorrelation or some other type of data structure assumption, the bootstrap demands resample in step 1 might be conducted differently. For example, we might bootstrap weeks or months of demands and randomly assemble the weeks/months into a new demand stream. Similarly, with autocorrelation we also might choose contiguous blocks of demands to calculate the LTD in step 2. Other techniques can also be employed to account for data structure; see Efron and

Tibshirani (1993, chapter 8) for additional discussion.

Specifically, under these more complicated scenarios, steps 1 and 2 *do not* reduce to the simple Monte Carlo approach previously described. Further, if the empirical lead time distribution is comprised of limited data, one might also add a step in between 1 and 2 to bootstrap the lead time distribution. Thus, it is useful to think about these types of problems in the more general bootstrap framework. This framework helps clarify the problem by decomposing into two sub-problems: (1) incorporating the uncertainty due to limited observed data (either demand, or lead time, or both) using bootstrap resampling, and (2) estimating the lead time demand distribution via Monte Carlo or some other appropriate means.

#### EVALUATION OF REORDER POINT PERFORMANCE WITH ACTUAL MARINE CORPS DATA

Ultimately, application of a new methodology to management of an inventory system should be evaluated with respect to the system's standard measures of performance. Prior to conducting that evaluation, we examine the current methodology for setting reorder points first on its own and then in light of alternative methodologies, focusing on service level performance directly attributable to the reorder point calculation. Three principal steps appear in our evaluation. We develop and apply a conservative upper bound that identifies ROPs actually in effect that are unreasonably large by even the most conservative assessment. We then compare the bootstrap technique, as developed in the previous section, with two other methodologies based on performance against several service targets. Finally, we perform a cost and service evaluation of bootstrap

ROPs in comparison with actual ROPs established at the I MEF General Account.

#### An Upper Bound for the ROP Requiring Minimal Information

When on-hand inventory reaches the ROP, the probability a stock out occurs,  $\alpha$ , is the probability that LTD exceeds the ROP. If we denote LTD by  $X$ , lead time (in days) by  $L$ , and demand (in items per day) by  $D$ , and assume that  $L$  and  $D$  are independent, then mean LTD is  $\mu_X = \mu_D \mu_L$  providing daily demands are also independent and share a common mean. Further, assuming demands are nonnegative, i.e., that quantities returned to the General Account are negligible,  $X$  is also a nonnegative random variable. We can therefore apply Markov's Inequality (see, e.g., Ross, 1993) directly, to obtain that

$$\alpha = P\{X > ROP\} \leq \frac{\mu_X}{ROP}, \quad (1)$$

or

$$ROP \leq \frac{\mu_X}{\alpha}. \quad (2)$$

Ouyang and Wu (1997) provide a tighter bound when a good estimate of LTD variance is available.

Provided demand is nonnegative and its mean constant over time, the reorder point should not exceed the average demand per period times average lead time divided by the desired maximum stock-out probability. The nonnegative demand assumption may not be warranted if an item experiences returns from users, but ignoring these returns is consistent with providing a conservative upper bound on the reorder point. Further, the bound holds for any distribution of demand and lead time; variance estimates, unreliable for lead time and poorly understood by Marines, are not required. The bound is very weak, of course, especially for low values of  $\alpha$ ; but it is easy

to calculate and can be used to highlight unreasonably high reorder points.

We apply (2) to identify such reorder points in the I MEF General Account, using lead time data from July 1997 to June 1998 and demand data from February 1997 to January 1998. To arrive at an estimate of mean lead time, we add an assumed 7-day per-order processing time to a reported average OST of 26.1 days (exclusive of items experiencing higher echelon backorders), totaling roughly 33 days. Specifying a risk (service) level of one (99) percent, we calculated the upper bound for each item's ROP using (2) and compared it to ROPs in effect on 27 April 1998 to check for unusually high actual ROPs. This resulted in 96 actual ROPs that were above the calculated upper bound. In fact, 44 items were between the upper bound and 150 percent of the upper bound, 26 were between 151 percent and 250 percent, and the other 26 were greater than 250 percent of the upper bound (the most extreme at 55 times the upper bound). The potential cost in inventory for these items, defined as the difference between the upper bound and the ROP, is more than \$350,000. That figure could be much larger if the actual ROP and on-hand stock could be set substantially lower than the upper bound. In any case, it is clear that there is applicability of the ROP inequality in practice and the ROPs for these 96 items deserve further scrutiny.

One cautionary note is in order when applying this bound. The ROP inequality, (2), is based on the means of the lead time and demand distributions, which will be unknown in practice. In applying (2) the means  $\bar{\mathcal{L}}$  and  $\bar{\mathcal{D}}$  have to be estimated based on averages taken from data, so the accuracy of the resulting bound will be a function of (a) the accuracy of the estimation of the means and (b) the adequacy with which the assumptions, used in the

derivation of the bound, are met. Thus, the accuracy of the bound must also be considered when evaluating whether an ROP is too large.

#### Comparison of Bootstrap, Order-Statistic, and Normal Approximation Techniques

The literature suggests that some techniques, such as the normal approximation for LTD, may be inappropriate for Marine Corps inventory systems. Another distribution-free approach, the order-statistic method of Lordahl and Bookbinder (1994), provides an alternative to the standard parametric techniques. Our primary basis for comparing these methodologies and the bootstrap technique is the ability of each to develop reorder points that can attain a specified service target with relative accuracy and precision.

Lordahl and Bookbinder (1994) estimate the  $100q^{\text{th}}$  quantile of LTD as follows. For a LTD sample of size  $n$ , find  $j+w=(n+1)q$ , where  $j = \lfloor (n+1)q \rfloor$ , the largest integer less than  $(n+1)q$ , and  $0 \leq w < 1$ . The reorder point providing an expected service level of  $q$  is then the weighted sum  $(1-w)X_{(j)} + wX_{(j+1)}$ , which is a simple linear interpolation of the two observations whose empirical CDF percentiles bracket  $q$ . Alternatively, under the assumption that the LTDs are normally distributed, Tersine (1994) selects the  $100q^{\text{th}}$  quantile from the normal distribution function best fitting the observed LTDs. In our implementation of these methods, we expressed any fractional reorder points to the next higher integer and followed the procedures previously established for determining reorder points from bootstrapped LTD distributions.



For the basis of comparison, we obtained daily demand data corresponding to

191 items with relatively high demand: for

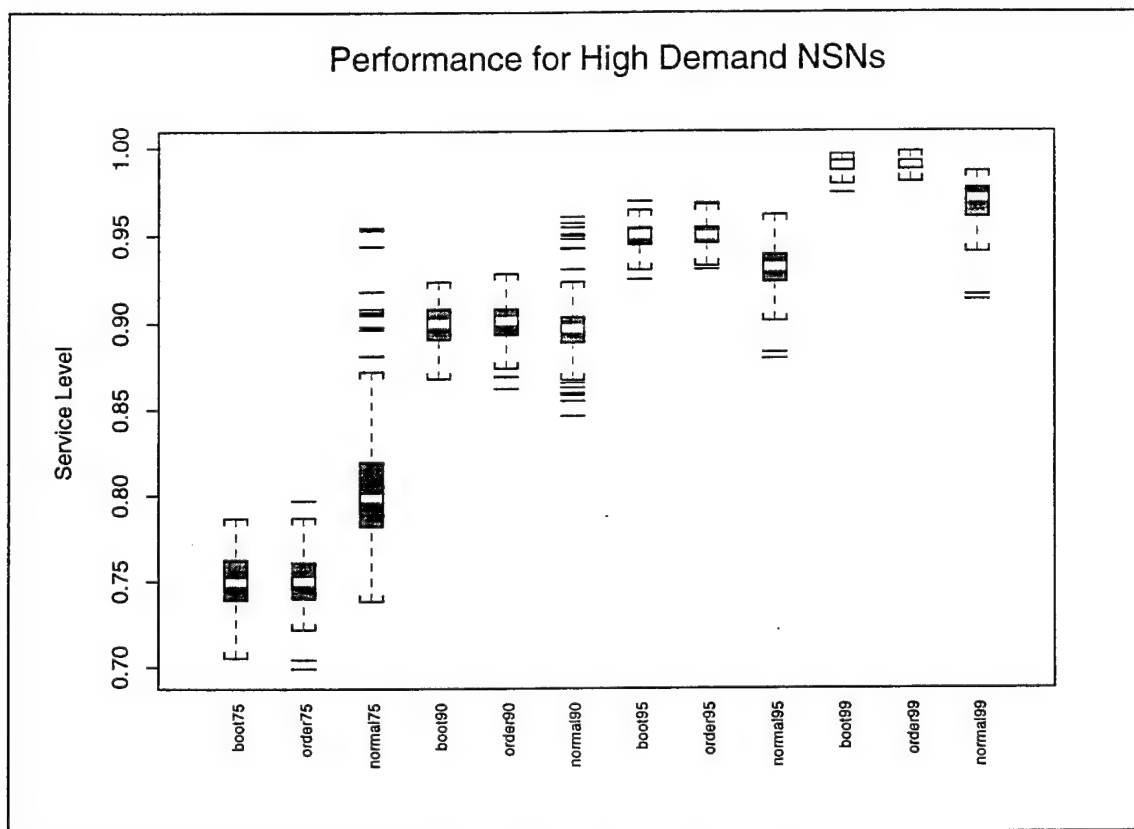


Figure 1. Comparative performance of the bootstrap (bootXX), order statistic (orderXX), and normal (normalXX) approximation methods of computing reorder points. Box plots show the mean performances of 191 items when reorder points are calculated according to the specified method for target service level "XX". While the bootstrap and order statistic techniques are on target (and virtually indistinguishable in terms of performance), the normal approximation method deviated significantly from established performance targets. It overstocked at the 75% target service level and failed to meet the 95% and 99% service level targets for more than half the items in each case (as shown by the interquartile range).

each item selected, annual quantity demanded exceeded 200 items on more than 100 requisitions. One year (February 1997 to January 1998) of daily demand data from I MEF were used. To determine reorder points for all three methods, we generated a single LTD sample of 1,000 observations by assuming a fixed 15-day lead time (approximately the median 1998 OST) and drawing at random, with replacement, 15 days' demands to total for each observation.

The bootstrap, order statistic and normal approximation methods all used these same 1,000 observations in determination of reorder points targeted at service levels of 75%, 90%, 95% and 99%. For each service level and reorder point methodology, we generated 100 new LTD observations using the same demand data and computed the frequency with which stock outs were observed. We repeated this experiment 30 times per item and calculated the mean

service level attained for each item; the results appear in Figure 1.

Figure 1 shows that, even when evaluated against the same data that produced them, reorder points determined by the traditional normal approximation are not robust enough for application to Marine Corps retail inventories such as the General Account. We calculated 95% confidence intervals for mean service level attained by each technique at each service level target; the confidence intervals for normal approximation performance included the intended target only at 90% service. Confidence intervals for bootstrap and order-statistic ROP performance always contained the intended targets. We similarly examined a sample of 100 items with low demand with comparable results. Further, as Figure 1 suggests, ROPs developed by the normal approximation technique varied widely in performance from item to item. A one-sided, approximate F test of the normal-to-bootstrap performance variance ratio rejected the null hypothesis that the ratio was one in every case at any reasonable level of significance. In fact, we should expect the item-to-item variance in service level performance to be higher for the normal approximation than for the distribution-free techniques, since the degree to which the approximation suitably represents LTD varies by distribution of LTD for each item.

The distinction between our methodology and the order statistic method of Lordahl and Bookbinder deserves remark. We determine reorder points from the cumulative distribution function of bootstrapped LTD observations, selecting the closest actual observation from the data instead of interpolating. The order statistic method interpolates between order statistics of actually observed LTDs. The principal difference, then, between the two methods is not in selection of a quantile, but in the use

of the bootstrap to obtain a LTD distribution. In this example, because the LTD distribution was simulated using resampling the only difference between the two procedures was in quantile selection. Thus, the performance difference between the two techniques was small. In practice, however, there could be greater differences in performance, particularly when the observed LTDs are few (often only one), such as with the Marine Corps data.

### A Test of Bootstrap ROPs

The preceding comparison of methodologies is a fair means of evaluating methodologies relative to one another, because it uses the same underlying empirical LTD distribution to evaluate service performance as it does to develop reorder points. This section shows how our bootstrap ROPs perform – in terms of service and cost – in a more realistic setting, when they are developed with historical demand data and evaluated against subsequent (actual) demands. We compare these reorder points against those determined by the Marine Corps for the same items, and present a comparison of number of lines (different types of items) stocked and inventory dollar value between the actual I MEF General Account and alternatives to it as developed with the bootstrap technique.

Table 1 shows the results of a controlled comparison of desired and achieved service levels for a fixed 30-day LTD and five service levels. We determined the bootstrap ROPs from data collected in the 12 months preceding February 1998 and evaluated them in five 30-day periods taken from calendar year 1998. Table 1 displays the desired service level against the proportion of items whose 30-day LTD did not exceed the ROP. For a given risk level and perfectly set ROPs, the expected number

of items experiencing a stock out should be the risk level multiplied by the total number

of items.

Service Level: (1- $\alpha$ ) x 100%	Percent of Lines that Did Not Stock Out for Five 30 Day Periods (by Date)				
	1-30 March	31 March – 29 April	30 April – 29 May	30 May – 28 June	29 June – 28 July
80.0	85.7	69.6	85.5	88.1	84.3
85.0	89.2	77.5	89.0	91.4	88.8
90.0	92.4	85.0	92.6	94.1	92.7
95.0	95.4	91.0	95.3	96.4	95.6
99.0	97.6	96.0	97.7	98.3	98.2

Table 1. Comparison of the probability of not stocking out (1- $\alpha$ ) to the actual observed rate for five thirty day periods for all lines (different types of items) that had a positive (nonzero) bootstrap ROP.

Table 1 shows that the 30 day fixed lead time bootstrap ROPs for a given service level track well when compared with the actual demands observed for five 30 day periods. Some bootstrap ROPs were zero; the table does not include those items. For all but one of the columns the actual results are slightly conservative for the smaller service level values, most likely as a result of the decision rule for selecting the ROP from the empirical distribution. However, the second column (31 March – 29 April) does show a period in which the demands were unusually high in comparison to the past year's demands (upon which the bootstrap calculations were based). This result demonstrates that any methodology, no matter how good, cannot perfectly predict the future.

Also note in Table 1 that, as the risk becomes very small, roughly  $\alpha < 0.05$ , the percentage of items not stocking out falls below the specified service level. This is because at very low risk levels around  $\alpha = 0.01$  the bootstrap essentially reduces to the sum of the maximum 30 days of demands observed in the past year and, again because the past is not always a good

indicator for the future, some items experience demands in the new period greater than the previous year's maximum. However, as the table shows, at small risk levels the difference between desired and actual performance is quite small—usually within a percentage point or two.

Overall, Table 1 demonstrates that our bootstrap methodology using independent sampling by days works well, from a practical standpoint, for these data. In order to get some insight into the variability of the ROP estimates, we bootstrapped standard errors of the estimated ROPs for the 191 items with relatively high demand selected in the "Comparison of Bootstrap, Order-Statistic, and Normal Approximation Techniques" subsection. For various service levels, we found that increasing the number of resamples from 1,000 to 10,000 decreased the standard errors by about a third. In particular, at  $\alpha = 0.05$  the average coefficient of variation for the 191 items decreased from around 11 percent to about four percent. However, increasing the number of resamples per item from 1,000 to 10,000 did not affect the aggregate results shown in Table 1 in any appreciable way, nor did

sampling by weeks instead of days. We therefore chose to use 1,000 resamples in our ROP calculations.

After evaluating service level performance with fixed lead time, we calculated a new set of ROPs using the bootstrap methodology for a variable lead time as follows. One year's worth of I MEF requisition data (1 February 1997 to 31 January 1998) was compiled into daily demands by item. That is, a matrix listing each item by total quantity requisitioned daily was created that showed how many units of each item were ordered on a day-by-day basis from the I MEF General Account. We fit a common log-normal distribution to aggregate lead time observations for replenishment orders. We then calculated the bootstrap ROPs (for variable lead times) as described in the "Using the Bootstrap to Calculate Reorder Points" section, performing the resampling using SAS (1990).

We defined the requisitioning objective, for those items with  $ROP_i > 0$ , as  $ROP_i + \max\{1, \text{round}(Z \text{ DOS}_i)\}$ , where  $\text{DOS}_i$  is the average demand for item  $i$  in terms of days of supply and the *round* function rounds the quantity to the next higher whole integer if the decimal fraction was greater than or equal to one-half, and it rounded down otherwise. (If  $ROP_i = 0$  then the requisitioning objective for item  $i$  is also zero.) That is, the operating level was set to a fixed days of supply,  $Z$ . For the bootstrap requisitioning objectives and ROPs, the on-hand quantity was randomly set uniformly between the ROP and requisitioning objective for each item. We refer to the bootstrap requisitioning objectives, ROPs, and derived on-hand quantities as the *bootstrap GABF*.

Table 2 compares the value of an inventory based on these new bootstrap GABFs to the actual value of the 27 April

GABF. In the third column, the table shows the results from the bootstrap GABF with a risk ( $\alpha$ ) equal to 0.05 and an operating level (OL) of 30 days of supply. The results show that roughly the same number of lines (different types of items) are stocked compared to the actual GABF, with a value at the ROP of less than half the actual GABF's value at the ROP. The other bootstrap GABF options illustrate how the number of lines and the dollar value of the requisitioning objective and ROP inventory increase for various values of decreasing risk.

Figure 2 compares the actual and bootstrap ROPs for the 35,068 items that had a positive ROP under either the actual or the  $\alpha=0.01$  bootstrap methodologies. Of the 35,068 items, 21,857 had a positive (nonzero) bootstrap ROP but were not actually stocked (i.e., the actual ROP was zero); 2,423 items were not stocked by the bootstrap methodology but had positive actual ROPs; and, 10,758 items had both positive actual and bootstrap ROPs. Visually there seems to be some consistency in the center of the graph, though the log-log scale tends to mask that small differences off the plot diagonal may be large actual differences. Also, over-striking plotting symbols on the two axes visually disguise the fact that that ten times as many items were not stocked by the actual methodology as compared to the bootstrap methodology and that twice as many points are plotted on the axes as in the center of the graph. That is, the center of the graph contains 10,758 points, and the vertical axis at bootstrap  $ROP=0$  has 2,423 points, while the horizontal axis at actual  $ROP=0$  has 21,857 points.

Of those items stocked by both methodologies, 46% of the time the bootstrap ROP is greater than the actual ROP and, conversely, 39 percent of the time

the actual ROP is greater than the bootstrap ROP. However, note that 77 percent of the complete list of 35,068 items had smaller actual ROPs because so many of them were simply not stocked using the current

methodology. Also, note that the value at the ROP for these two options is virtually the same; thus, the bootstrap ROP stocks 2-1/2 times as many types of items for the same cost. Under the assumption that the

	Actual GABF	Bootstrap GABFs			
		$\alpha=0.05$ OL=30	$\alpha=0.03$ OL=30	$\alpha=0.01$ OL=30	$\alpha=0.01$ OL=90
Number of lines stocked	13,159	14,014	17,682	32,645	32,645
Value at the ROP	\$13.5M	\$4.7M	\$6.7M	\$13.2M	\$13.2M
Value on-hand	\$25.5M	\$5.1M	\$7.2M	\$13.7M	\$21.3M
Value at the requisitioning objective	\$23.6M	\$6.4M	\$8.8M	\$16.9M	\$24.5M

Table 2. Comparison of the actual GABF (27 April 1998) to various alternative bootstrap GABFs. The operating level, OL, is specified in terms of days of supply.

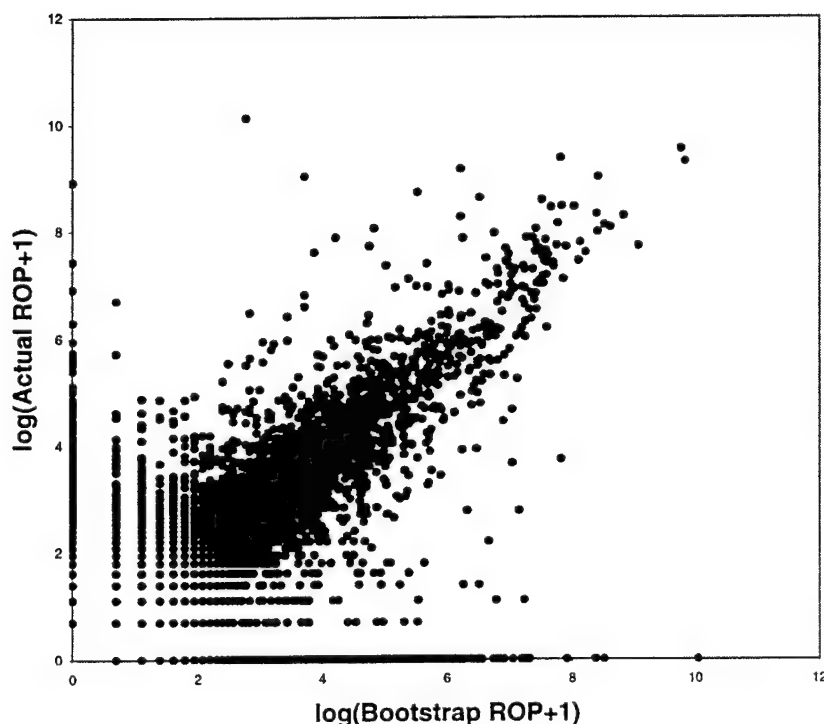


Figure 2. Comparison of the actual 27 April GABF ROPs to the Bootstrap ROPs.



bootstrap ROPs better reflect the variability of demand, it seems clear that the existing methodology must unnecessarily overstock expensive items.

### Examples of Bootstrap ROPs

In order to demonstrate how the bootstrap ROP works it is instructive to look at some specific cases. Table 3 shows four items with varying demand patterns along with three possible bootstrap ROPs corresponding to risks of 0.01, 0.05, and 0.10. The first two cases had a total of five requisitions for the year from February 1997 to January 1998, with varying total quantities requested. The last two cases have a larger number of requisitions.

Case 1 is an item that had five requisitions with a total of 488 items requested for the 12 month period. For the three different risk levels considered,  $\alpha=0.10, 0.05$ , and  $0.01$ , the bootstrap ROPs are 50, 108, and 282, respectively. Note that, as the allowable risk is decreased (equivalently, the service level is increased), the ROP increases. Case 2 shows a different situation in which 5 expensive items (\$1,652.00 each), each on a different requisition, all occurred in a short span of time. In this case, the bootstrap ROPs for the higher risk levels set the ROP to zero, but for the smallest risk (highest service level) it is set to five units. Here the methodology performs according to intuition: It either sets the ROP to cover the entire batch of 5 or not to cover them at all, depending on the risk that the inventory manager is willing to assume.

Cases 3 and 4 show items with higher requisition activity. Case 3 shows that the largest bootstrap ROP can be smaller than the greatest month (March) if the requisitions within that month are

sufficiently spread out over time. Case 4 has a high number of requisitions but a low number of units per requisition. Taken together, the four cases in Table 3 cover a wide spectrum of demand patterns and unit prices. All of these cases show that the bootstrap methodology sets reasonable ROPs that correspond with intuition, and they illustrate how an inventory manager can specify a risk level and the methodology will set appropriate ROPs.

Contrast the bootstrap ROP results to the current USMC methodology, as shown in the "USMC ROP" and "USMC RO" columns (the latter denoting requisitioning objective). For every case, not only is the ROP insufficient, as illustrated by the fact that most cases have multiple months for which the ROP would not cover demand, but even the requisitioning objective is insufficient for at least one month in every case. Indeed, the USMC methodology, for these four cases, looks similar to the  $\alpha=0.1$  bootstrap case, which means that roughly ten percent of the items should stock out when they are at the ROP. Multiply this rate times thousands of items and the number of stock outs is large.

### EVALUATING THE OVERALL EFFECT OF THE BOOTSTRAP ROPs

Evaluating the bootstrap ROPs in relation to the existing ROPs gives little indication of how a new bootstrap GABF, based on the new ROPs, would actually perform in the real world in terms of meaningful inventory metrics. To evaluate this question, a more detailed set of experiments was conducted in which the performance of the entire General Account was simulated, starting at an initial inventory position and then "playing back" all the

demands for eight months while measuring the performance of the inventory.

Case	NSN (Item)	Unit Price (\$)	USMC ROP	USMC RO	Total Annual		Bootstrap ROP			1997												1998	
					Quantity Requested	# of Req's	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan		
1	8415010333520	0.06	47	127	488	5	50	108	282	0	0	282	0	60	48	98	0	0	0	0	0	0	0
2	2940010882429	1652.00	1	2	5	5	0	0	5	0	0	0	5	0	0	0	0	0	0	0	0	0	0
3	1005007266131	808.00	9	19	63	50	9	13	18	0	26	3	5	1	1	2	2	1	5	13	4	4	4
4	5325011917555	4.57	17	46	174	104	12	24	74	7	13	24	6	25	5	0	2	0	3	77	12	12	12

Table 3. Various items and their annual demand pattern for one year with three possible bootstrap ROPs corresponding to risks of 0.01, 0.05, and 0.10. "RO" denotes requisitioning objective.

In performing such a "real world" experiment, we did not seek specifically to obtain results of an analytical nature. Instead, we compared the performance of a simulated system with that of a real inventory (the I MEF General Account) over a specific period. In order to make the comparison meaningful to inventory managers at the I MEF General Account, we had to depict how the simulated system responded to known, identifiable events that occurred during the specific period. In this way, both the simulated system and the real inventory were subject to the same resource constraints and had the same sequence of problems to solve. Collection of summary statistics from the simulation enabled comparison of actual and simulated metrics over the period of interest. Determination of reorder points for each item in the simulated system relied on empirical demand and lead time distributions collected prior to the period of interest, and proceeded according to the procedures we have explained.

Such an experiment required making assumptions for transaction information that was not available, such as the exact order that daily demands arrived at the General Account. It also required simulation of the stock replenishment process. However, given that such assumptions were equally applied to all experiments, and that the performance of the system using an actual inventory starting position results in system performance (as measured by inventory metrics such as those described in "Performance Metrics" subsection) that mirrors real world system performance, one would expect that bootstrap GABF improvements shown in the simulation would portend similar improvements in practice.

In general, the experiments were conducted as follows:

1. Begin with an initial inventory position from either the actual GABF (27 April 1998) or a bootstrap GABF (calculated using February 1997 through January 1998 demand data). Set the on-hand stock at the actual on-hand stock for the existing GABF or, for the bootstrap GABF, uniformly between the requisitioning objective and ROP for each item and start with a "clean slate" with zero due-in and due-out for all items.
2. Take eight months' of "new" demand data from 1 May through 31 December 1998 and play these back day-by-day, subtracting each demand from the on-hand stock if available or adding to the due-out if not (or "pass through" to the next higher supply echelon in accordance with standard Marine Corps practices).
3. Track the inventory position for all items over time. Weekly, rank those items with inventory position less than or equal to the ROP and place replenishment buys subject to a fiscal constraint equal to that normally experienced in practice. Stochastically assign due-in dates for the replenishment requisitions and add them in to the on-hand stock when they come in.
4. Maintain inventory statistics over the course of the simulation and calculate overall performance metrics at the conclusion of the quarter.

We specifically refer to these exercises as experiments and not simulations because the demand stream was actual data, as was the 27 April GABF. Simulation only entered into the experiment to account for unobserved actions or unavailable data, such

	Actual GABF	Bootstrap GABFs			
		$\alpha=0.05$ OL=30	$\alpha=0.03$ OL=30	$\alpha=0.01$ OL=30	$\alpha=0.01$ OL=90
ERO fill rate	53.9%	52.0%	58.2%	68.4%	66.3%
Fill rate	72.2	72.7	77.5	85.1	84.8
Satisfaction rate	81.0	86.9	89.7	92.3	91.9
Accommodation rate	83.1	83.7	86.3	92.2	92.2

Table 4. Comparison of the actual GABF performance compared to various bootstrap GABFs. The operating level, OL, is specified in days of supply.

as the order which the demands were presented to the General Account or stock-outs in the higher echelons of the supply system. As much as possible we sought simply to replay what actually occurred. Indeed, in the experiments using the actual GABF, the resulting inventory metrics were found to mirror those observed in actual practice.

Table 4 shows that using the bootstrap GABF with  $\alpha=0.01$  increased the fill rate from 72.2 percent (using the original GABF) to about 85 percent and the ERO fill rate increased from 53.9 percent to around 68 percent. These results were achieved, as indicated in Table 2, with an on-hand inventory value substantially less than the existing inventory value, where the on-hand inventory value for the  $\alpha=0.01$  (OL=30) bootstrap GABF was under \$14 million as compared to the actual on-hand inventory position of almost \$26 million. Perhaps the most striking contrast comes with the comparison of the  $\alpha=0.05$  bootstrap GABF to the actual GABF. This bootstrap GABF showed roughly equivalent performance in

all of the metrics in Table 4 using an inventory with a fraction of the dollar value of the actual GABF (Table 2). Essentially, this bootstrap GABF achieved marginally better performance for less than half the price of the actual inventory.

Table 4 also shows the performance improvements that can be achieved by the bootstrap GABFs for various risk levels, where we see that as the risk is reduced the various metrics increase accordingly. For example, the  $\alpha=0.01$  bootstrap GABFs have an accommodation rate of 92 percent, which means that 92 percent of all items requested had a positive requisitioning objective (in contrast to 83 percent in the actual GABF). Similarly, 92 percent of the demands for items with a positive requisitioning objective were actually filled out of local stock (the satisfaction rate) with the  $\alpha=0.01$  bootstrap GABFs, while only 81 percent were with the actual GABF. Even larger differences result for the fill rates and the ERO fill rates. The last two columns also demonstrate that the bootstrap ROPs actually do account for almost all of



the demand variability in the safety level because the various "rates" change insignificantly when the operating level is tripled from 30 days of supply to 90 days of supply.

We have also found that inventory performance can be improved further by targeting various classes of parts by dollar value (expensive) and appropriately allowing an increased risk level. That is, using a risk level of 1 percent for all parts is to treat everything as highly important when, in fact, there are certainly groups of parts that could be allowed a higher risk with little detriment to the inventory's performance. The inventory costs thus avoided by assuming a slightly greater risk for a small class of expensive parts can then be invested in substantially decreasing the risk for a large class of inexpensive parts. We have applied such "dollar banding" techniques to this problem and, for the same inventory investment costs (\$13 million at the ROP, \$17 million at the requisitioning objective), have achieved a 90 percent fill rate (with a 98 percent accommodation rate and a 92 percent satisfaction rate) and a 76% ERO fill rate.

Of course, the adoption of simple cost-based rules in a military system may not always be appropriate. The major point, however, is that risk of stock out may be assigned to classes of parts using the bootstrap methodology. Price classification is only one such approach; military importance might also be used. Additionally, the current use of a fixed 60 days of supply for the operating level, and perhaps even the 30 days of supply used in some of the bootstrap GABFs, is excessively large. ROPs with appropriately sized safety levels, as with the bootstrap ROPs, should allow the inventory to function with smaller operating level yet with a higher confidence of fewer stock-outs. Smaller operating

levels require smaller incremental outlays of funding to buy items from the ROP back to the requisitioning objective, and result in other ancillary benefits such as smoothing work flow in the warehouse.

## SUMMARY

This paper has presented a nonparametric bootstrap methodology for setting inventory reorder points and a simple inequality for identifying unreasonably high ROPs. We have shown that an empirically-based bootstrap method is feasible and calculable for a large inventory, such as that maintained by each Marine Expeditionary Force. For the Marines we have shown that the methodology should work significantly better than the existing methodology, and it is preferable to standard parametric procedures based on normality assumptions.

In the "Evaluation of Reorder Point Performance with Actual Marine Corps Data" section, we applied Markov's Inequality to identify 96 items with unreasonably high reorder points (in only one of the three active-force General Accounts) to illustrate the inadequacy of the existing methodology. Our approach of extracting quantiles of a bootstrapped LTD distribution resulted in no reorder points exceeding these upper bounds. Implementation of the bootstrap methodology further demonstrated the potential cost benefits of implementing an improved methodology for stocking the General Account.

It is worth noting that the bootstrap methodology generates a complete empirical distribution of LTD, from which all possible ROPs (for each item) can be related to stock out risks, useful for additional purposes. For example, under fiscal constraints the estimated LTD distribution could be used to rank items by "risk differential," meaning the difference between the risk at the current

on-hand stock position and the desired stock position (either requisitioning objective or ROP). Such a ranking can then be used to decide how to spend limited funds to increase the inventory position of those items most at risk of stock-out.

The bootstrap methodology can also be generalized to more complicated inventory problems, such as stocking repairable items. In this scenario, an item that is issued is coupled with a returned broken item that may be repaired or refurbished and returned to stock. Thus, stock replenishment comes from two sources, the receipt of replenishment orders for new items from the next higher supply echelon, and the receipt of refurbished items from a repair shop. The two sources of supply will have differing lead times, and the refurbished items are likely to arrive in some stochastic manner. Clearly this is a much more complicated scenario that is not likely to lend itself to simple analytic and parametric methods. Yet, with the appropriate data, a bootstrap scheme can be readily implemented which would account for these additional requirements and still estimate a LTD distribution from which ROPs could be set in the same manner as the other inventory problems presented in this paper.

Future work will involve finding more sophisticated methods for setting the RO and accounting for the interaction between the RO and the ROP. In particular, because the risk being fixed in this methodology is conditional on the inventory position being at the ROP, the risk may be underestimated for items typically ordered in sizeable batches. This is because the inventory position for such items could have a tendency to overshoot the ROP, causing it to be substantially lower than the ROP when the replenishment order is placed. Also, in keeping with most other approaches in the literature, the bootstrap methodology does

not calculate an overall risk of stock-out—arguably the quantity of most interest to an inventory manager—which is both a function of the ROP and of the operating level. However, our results here have shown that, within the bounds of reasonable operating level choices, such issues are second order considerations and the Marine Corps would achieve significant inventory performance improvement using the bootstrap scheme we have proposed.

The major requirements for the bootstrap approach are (a) historical data on which to run the bootstrap and (b), for large inventories, sufficient computational power. However, we have shown that the former requirement can be met with one year's historical data and the latter, for MEF-sized data, can be met with the processing power of a high-end personal computer today. Both of these problems are easily solved with time, as more data are collected and as personal computers become more powerful. In such an environment, the use of empirically based methods such as those presented here become increasingly attractive.

#### ACKNOWLEDGEMENTS

Lieutenant Colonel James Kessler and his staff at 1<sup>st</sup> Force Service Support Group, I MEF, including Major Terry Flannery and Chief Warrant Officer Barry Newland, were particularly helpful in providing data and discussing supply operations in the Marine Corps. Without their assistance, none of this work would have been possible. Marc Robbins provided invaluable guidance and insight in the preparation and execution of this work, and Mark Totten and Pat Boren performed most of the SAS programming that underlies the summary statistics contained in this paper. Finally, Dan McCaffrey and Brad Efron were generous

with their time and thoughts as we worked through these ideas.

## REFERENCES

- Arrow, K. J., Karlin, S. and Scarf, H. 1958 *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, California.
- Azoury, K. S. and Brill, P. H. 1992 "Analysis of net inventory in continuous review models with random lead time," *European Journal of Operational Research* **59**, 383-392.
- Bookbinder, J. H. and Lordahl, A. E. 1989 "Estimation of Inventory Re-Order Levels Using the Bootstrap Statistical Procedure," *IIE Transactions* **21**(4), 302-312.
- Efron, B. and Tibshirani, R. J. 1993 *An Introduction to the Bootstrap*, Chapman & Hall, Inc., New York.
- Eppen, G. D. and Martin, R. K. 1988 "Determining Safety Stock in the Presence of Stochastic Lead Time and Demand," *Management Science* **34**(11), 1380-1390.
- Fricker, R. D., Jr., and Robbins, M. 1999 "Retooling for the Logistics Revolution: Designing Marine Corps Inventories to Support the Warfighter," RAND, MR-1096-USMC.
- Harrell, F. E. and Davis, C. E. 1982 "A new distribution-free quantile estimator," *Biometrika* **69**(3), 635-640.
- Gallego, G. 1992 "A minmax distribution free procedure for the (Q, R) inventory model," *Operations Research Letters* **11**, 55-60.
- Gallego, G. and Moon, I. 1993 "The distribution free newsboy problem: review and extensions," *Journal of the Operational Research Society* **44**, 825-834.
- Lordahl, A. E. and Bookbinder, J. H. 1994 "Order-Statistic Calculation, Costs, and Service in an (s, Q) Inventory System," *Naval Research Logistics* **41**, 81-97.
- Moon, I. and Gallego, G. 1994 "Distribution Free Procedures for Some Inventory Models," *Journal of the Operational Research Society* **45**(6), 651-658.
- Ouyang, L. and Wu, K. 1997 "Mixture Inventory Model Involving Variable Lead Time with a Service Level Constraint," *Computers & Operations Research* **24**(9), 875-882.
- Ross, S. M. 1993 *Introduction to Probability Models* (3rd ed.), Academic Press, San Diego, California.
- SAS Institute Inc. 1990 *SAS® Procedures Guide* Version 6, SAS Institute Inc., Cary, North Carolina.
- Selen, W. J. and Wood, W. R. 1987 "Inventory Cost Definition in an EOQ Model Application," *Production and Inventory Management Journal* **28**(4).
- Tersine, R. J., *Principles of Inventory and Materials Management* (4th ed.), Prentice Hall, Englewood Cliffs, New Jersey, 1994.
- Tijms, H. and Groenevelt, H. 1984 "Simple approximations for the reorder point in periodic and continuous review (s, S) inventory systems with service level constraints," *European Journal of Operational Research* **17**, 175-190.
- Tyworth, J. E. and O'Neill, L. 1997 "Robustness of the Normal Approximation of Lead-Time Demand in a Distribution Setting," *Naval Research Logistics* **44**, 165-185.
- United States Marine Corps 1992 *Intermediate-Level Supply Management Policy Manual*, MCO P4400.151B.

Wang, M. C. and Rao, S. S. 1992

“Estimating reorder points and other management science applications by bootstrap procedure,” *European Journal of Operational Research* **56**, 332-342.

#### ACRONYMS

---

ERO	Equipment Repair Order
GABF	General Account Balance File
LTD	Lead Time Demand
MEF	Marine Expeditionary Force
NSN	National Stock Number
OL	Operating Level
OST	Order and Ship Time
RO	Requisitioning Objective
ROP	Reorder Point
USMC	United States Marine Corps

---

### DESCRIPTORS

Logistics, Nonparametric, Inventory, Lead Time Demand, Supply Systems, Markov's Inequality, Bootstrap, Military, United States Marine Corps